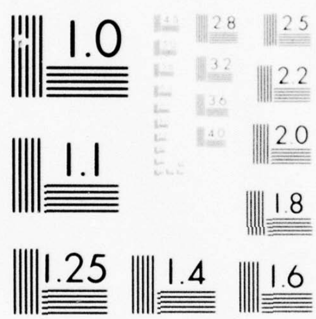
AD-A045 166 TEMPLE UNIV PHILADELPHIA PA DEPT OF MATHEMATICS F/G 12/1
ON MEASURES OF DEPENDENCE. A SURVEY OF RECENT DEVELOPMENTS. REV--ETC(U)
JUL 77 S KOTZ, C SOONG AFOSR-75-2837
UNCLASSIFIED AFOSR-TR-77-0745-REV NL

| OF |
AD
A045166

END
DATE
FILMED
11 -77
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

On Measures of Dependence

A survey of recent developments

S. Kotz and C. Soong

Dept. of Mathematics
Temple University
Philadelphia, Pa. 19122

July 1977

Revised version

Research supported by the
U. S. Air Force
Office of Scientific Research,
Grant No. 75-2837

Abstract

The purpose of these notes is to survey the recent literature on dependence between two random variables and to stimulate research which will extend (some of) these concepts and relations to the case of 3 or more variables.

Key words: bivariate distributions; FGM distributions; multivariate dependence; Hoeffding's lemma; multivariate hazard rates; correlation ratio; copulas; association; information theory; reliability.

Part 1 Dependence and generalizations of correlation.

I) Introduction.

In these notes we are concerned mainly with bivariate distributions. The connection between dependence and the theory of multivariate hazard rates is also briefly discussed. Extensions of positive dependence to association and other notions of dependence as well as applications and interrelations are included in Part 2 of these notes.

II) Positive dependence. (Lehmann, 1966).

Two *events* A and B are called dependent if $P(A \cap B) = P(A)P(B)$ is *violated*. We say that there exists positive dependence if

$$P(A \cap B) \geq P(A)P(B).$$

Two random variables are said to be *positively dependent* if $P(X \in A,\ Y \in B) \geq P(X \in A)P(X \in B)$ for any two Borel sets A and B on the real line. Negative dependence is defined by reversing the appropriate inequalities. The former case often occurs in reliability theory (parts of a machine usually have a longer life when they are "put together"). The latter case is prevalent in biological populations competing for limited resources. (See Barlow and Proschan (1975).)

From Lehmann (1966) we have the following definitions:

Def. 1. (X,Y) or $F_{XY}(x,y)$ is positively quadrant dependent if:

$$P(X\le x,\ Y\le y) \ge P(X\le x)P(Y\le y) \quad \text{or}$$
$$F_{XY}(x,y) \ge F_X(x)F_Y(y) \quad \forall \ x,y. \tag{1.1}$$

Let F_1 be the family of distributions (two-dimensional) for which (1.1) is valid, G_1 be the family of distributions for which (1.1) is valid with reversed inequality. Notation $(X,Y) \in F_1(G_1)$ means $F_{XY} \in F_1(F_{XY} \in G_1)$.

<u>Lemma 1</u>: (i) $(X,X) \in F_1$

(ii) $(X,Y) \in F_1 \iff (X,-Y) \in G_1$

(iii) $P(X\le x,\ Y\le y) \ge P(X\le x)P(Y\le y)\ \forall\ x,y \iff$

$P(X\le x,\ Y<y) \ge P(X\le x)P(Y<y) \quad \forall \ x,y \iff$

$P(X<x,\ Y<y) \ge P(X<x)P(Y<y) \quad \forall \ x,y.$

<u>Proof</u>: (i) $P(X\le x,\ X\le x) = P(X\le x) \ge P(X\le x)P(X\le x)$.

(ii) $\Rightarrow P(X\le x,\ -Y\le y) = P(X\le x,\ Y\le -y)$

$$\begin{aligned}
&= P(X\le x) - P(X\le x,\ Y<-y) = \\
&= P(X\le x) - \lim_{n\to\infty} P(x\le x,\ Y\le -y - \tfrac{1}{n}) \\
&\le P(X\le x) - \lim_{n\to\infty} P(X\le x) \cdot P(Y\le -y - \tfrac{1}{n}) \\
&= P(X\le x) - P(X\le x) \cdot P(Y<-y) \\
&= P(X\le x)\left[1 - P(Y\le -y)\right] = P(X\le x)P(-Y\le y).
\end{aligned}$$

for $\Leftarrow$ the proof is similar.

(iii) Since $P(X\le x,\ Y<y) = \lim_{n\to\infty} P(X\le x,\ Y\le y - \frac{1}{n})$ and also $P(X\le x,\ Y\le y) = \lim_{n\to\infty} P(X\le x,\ Y<y+\frac{1}{n})$, we proceed as in (ii).

It is easy to verify the following.

Lemma 2: $(X,Y) \in F_1$

$$\Longleftrightarrow P(X\le x,\ Y\ge y) \le P(X\le x)P(Y\ge y)\ \forall\ x,y$$

$$\Longleftrightarrow P(X\ge x,\ Y\le y) \le P(X\ge x)P(Y\le y)\ \forall\ x,y$$

$$\Longleftrightarrow P(X\ge x,\ Y\ge y) \ge P(X\ge x)P(Y\ge y)\ \forall\ x,y.$$

Remarks: 1. The signs $\ge$ and $\le$ can be replaced by $>$ and $<$ respectively.

2. The last inequality is called G-*dependence* by Lehmann (1966) and Johnson and Kotz (1975), and is frequently used in reliability theory where X and Y are interpreted as life-lengths of component parts of a machine. The validity of the last inequality follows from the simple relation:

$$F_{XY}(x,y) - F_X(x)F_Y(y) = G_{XY}(x,y) - G_X(x)G_Y(y)$$

where $G_{XY}(x,y) = P(X>x,\ Y>y)$ and $G_X(x) = P(X>x)$ and $G_Y(y) = P(Y>y)$.

We conclude this section by proving that $(X,Y) \in F_1 \Rightarrow E(XY) \ge EXEY$ provided the covariance and expectations exist.

Lemma 3: (Hoeffding 1940): If F denotes the joint and F_X and F_Y denote the marginal distributions of X and Y, then

$$E(XY) - EXEY = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [F(x,y) - F_X(x)F_Y(y)]dxdy =$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(G_{XY}(x,y) - G_X(x)G_Y(y)\right)dxdy,$$

provided the expectations on the l.h.s. all exist.

Proof: Define $I(u,x) = 1$ if $u<x$ and 0 otherwise.

Let (X_1,Y_1), (X_2,Y_2) be independent, each distributed according to F. Then

$$2\,\mathrm{Cov}(X_1Y_1) = 2\big(E(X_1Y_1) - EX_1EY_1\big) = E(X_1 - X_2)\cdot(Y_1 - Y_2) =$$

$$= E\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[I(u,X_1) - I(u,X_2)][(v,Y_1) - I(v,Y_2)]\,dudv.$$

(Details of the last step are given in Appendix 1.) Using Fubini's theorem we can take the expectation inside the integral sign thus obtaining:

$$2\,\mathrm{Cov}(X_1,Y_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\big(E\,I(u,X_1)I(v,Y_1) - E\,I(u,X_2)\cdot E(v,Y_1) -$$

$$E\,I(u,X_1)\,E\,I(v,Y_2) + E\,I(u,X_2)I(v,Y_2)\big)\,dudv$$

$$\left(= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{Cov}\big(I(u,X_1),\, I(v,Y_1)\big)\right)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\Big(G_{X_1,Y_1}(u,v) - G_{X_2}(u)G_{Y_1}(v) - G_{X_1}(u)G_{Y_2}(v)$$

$$+\, G_{X_2,Y_2}(u,v)\Big)\,dudv$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}2\Big(G_{X_1Y_1}(u,v) - G_{X_1}(u)\cdot G_{Y_1}(v)\Big)\,dudv.$$

Using the Remark 2 following Lemma 2 we complete the proof. □

Lemma 3 implies:

Theorem 1: If $(X,Y) \in F_1$ and EXY, EX, EY exist, then $EXY \ge EXEY$ or $\mathrm{Cov}(X,Y) \ge 0$. Moreover, if $(X,Y) \in F_1$, $\mathrm{Cov}(X,Y) = 0 \Rightarrow X,Y$ are independent.

Some additional properties of positive quadrant dependence.

Given:

(i) $$P[X\le x,\ Y\le y] \ge P[X\le x]P[Y\le y] \quad \forall \quad x,y \in R' .$$

The following is valid:

(ii): $$P[X>x,\ Y\le y] \le P[X>x]P[Y\le y]$$

(iii): $$P[X\le x,\ Y>y] \le P[X\le x]P[Y>y]$$

(iv): $$P[X>x,\ Y>y] \ge P[X>x]P[Y>y],$$ and

(v): (i) does not imply

$$P[x_1<X\le x_2,\ y_1<Y\le y_2] \ge P[x_1<X\le x_2]P[y_1<Y\le y_2] \qquad (-\infty<x_1,x_2,y_1,y_2<\infty).$$

Proofs:

(ii): $$\begin{aligned} P(X>x,\ Y\le y) &= P(Y\le y) - P(X\le x,\ Y\le y) \\ &\le P(Y\le y) - P(X\le x)P(Y\le y) \\ &= \big(1 - P(X\le x)\big)P(Y\le y) \\ &= P(X>x)P(Y\le y). \end{aligned}$$

(iii): Follows from (ii) by symmetry.

(iv): From (iii)

$$\begin{aligned} P(X>x,\ Y>y) &= P(Y>y) - P(X\le x,\ Y>y) \\ &\ge P(Y>y) - P(X\le x)P(Y>y) \\ &= \big(1 - P(X\le x)\big)P(Y>y) \\ &= P(X>x)P(Y>y) \end{aligned}$$

Note that (iv) is just the Remark 2 above.

(v): Counterexample: Let (X,Y) be distributed according to the following table.

Y \ X	1	2	3	
1	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{11}{18}$
2	0	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{5}{18}$
3	0	0	$\frac{1}{9}$	$\frac{2}{18}$
	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1

$P[X\le a,\ Y\le b] \ge P[X\le a]P[Y\le b]$ for $a=1,2,3$ and $b=1,2,3$.

$P[X\le 1,\ Y\le 1] = \frac{1}{3} \ge \frac{1}{3} \times \frac{11}{18}$

$P[X\le 2,\ Y\le 1] = \frac{1}{3} + \frac{1}{6} \ge \frac{2}{3}\left(\frac{11}{18}\right)$

$P[X\le 2,\ Y\le 2] = \frac{2}{3} \ge \frac{2}{3}\left(\frac{16}{18}\right)$

$P[X\le 1,\ Y\le 2] = \frac{1}{3} \ge \frac{1}{3} \times \frac{16}{18}$, and so on.

However $P[1.5\le X\le 2.5,\ .5\le Y\le 1.5] < \frac{1}{3}\cdot\frac{11}{18} = P[1.5\le X\le 2.5]\cdot P[.5<Y<1.5]$. Indeed, $P[X=a,\ Y=b]$ cannot always be $\ge P[X=a]P[Y=b]$ (excluding the independent case) since both $\sum_b\sum_a P[X=a,\ Y=b] = 1$ and $\sum_b\sum_a P[X=a]P[X=b] = 1$.

NOTE: Inequalities (i) and (ii) imply that there must be Borel sets A_1, B_1 such that $P[X\epsilon A_1,\ Y\epsilon B_1] \le P[X<A_1]P[Y\epsilon B_1]$ and sets A_2, B_2 such that $P[X\epsilon A_2,\ Y\epsilon B_2] \ge P[X\epsilon A_2]P[Y\epsilon B_2]$ (with a strict inequality for at least one pair). For example we may choose

$$A_1 = [X>x],\quad A_2 = [X\le x]$$

and $\quad B_1 = B_2 = [Y\le y]$.

III) Measures of dependence from information-theoretic aspects.

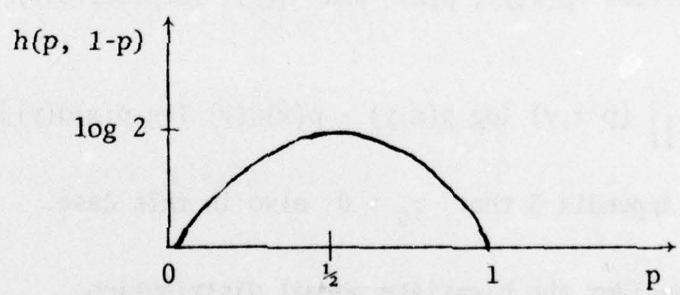
The following discussion develops a measure of dependence based on the concept of entropy suggested by C. E. Shannon (1948) nearly three decades ago.

Let X be a discrete random variable with a finite number of outcomes, i.e. $P(X=x_i) = p_i$, $i = 1,2,\ldots N$. Shannon defines the *entropy* or the measure of uncertainty (or information) as:

$$h(\underset{\sim}{p}) = -\sum_{i=1}^{N} p_i \log p_i, \quad \underset{\sim}{p} = (p_1, p_2, \ldots, p_N), \quad \sum_{i=1}^{N} p_i = 1.$$

$h(\underset{\sim}{p})$ assumes maximum $\log N$ iff $p_i = \frac{1}{N}$, $i = 1,\ldots N$.

The proof of this assertion as well as some other properties of $h(\underset{\sim}{p})$ are presented in Appendix 2. Here we take the particular case $N=2$. In this case:



$$h(p, 1-p) = -p \log p - (1-p)\log(1-p); \quad h(\tfrac{1}{2},\tfrac{1}{2}) = \log 2.$$

We define $h(0,1) = \lim_{p\to 0} h(p, 1-p) = 0$, $h(1,0) = \lim_{p\to 1} h(p, 1-p) = 0$. These limits exist by the L'Hospital rule. (Observe that an unbiased coin yields a higher uncertainty than a biased one.) Consider discrete r.v.'s X,Y with

$$P(X=x_i,\ Y=y_j) = p_{ij} \qquad (i = 1,\ldots n;\ j=1,\ldots m);$$

$P(X=x_i) = p_i$, $P(Y=y_j) = q_j$. We have

$$q_j = \sum_i p_{ij}\ ,\quad p_i = \sum_j p_{ij} \qquad \text{and}$$

by the lemma in Appendix 2,

$$-\sum p_i q_j \log(p_i q_j) \geq -\sum p_{ij} \log p_{ij}\ .$$

This inequality suggests that we have a higher uncertainty when X and Y are independent. This fact is employed to define the *logarithmic index* of correlation r_0 :

$$r_0 = \sum_{i,j} \left(p_{ij} \log p_{ij} - p_i q_j \log p_i q_j\right).$$

Note that $r_0 \geq 0$, and $r_0 = 0$ iff $p_{ij} = p_i q_j$ $\forall$ i,j. If (X,Y), X, Y possess densities $p(x,y)$, $p(x)$ and $q(y)$ respectively, we define r_0 as:

$$r_0 = \iint [p(x,y) \log p(x,y) - p(x)q(y) \log p(x)q(y)]dxdy.$$

We prove in Appendix 3 that $r_0 \geq 0$ also in this case.

<u>Example</u>: Consider the bivariate normal distribution:

$$p(x,y) = \frac{1}{2\pi} \sqrt{ab-h^2}\ e^{-\frac{1}{2}(ax^2+2hxy+by^2)}\ ,\quad a>0,\quad ab-h^2 > 0.$$

The classical correlation coefficient is:

$$r = \frac{EXY-EXEY}{\sqrt{Var\ X}\ \sqrt{Var\ Y}} = \frac{-h}{\sqrt{ab}}\ .$$

The marginals are

$$p(x) = \int_{-\infty}^{\infty} p(x,y)dy = \sqrt{\frac{\alpha}{\pi}}\, e^{-\alpha x^2} \quad ; \quad q(y) = \int_{-\infty}^{\infty} p(x,y)dx = \sqrt{\frac{\beta}{\pi}}\, e^{-\beta y^2} ,$$

$$\alpha = (ab - h^2)/2b \quad ; \quad \beta = (ab - h^2)/2a.$$

(See also the next Section.)

Computing r_0 , we obtain

$$r_0 = -\log \frac{2\pi e}{\sqrt{ab-h^2}} + \log\sqrt{\frac{\pi e}{\alpha}} + \log\sqrt{\frac{\pi e}{\beta}}$$

$$= \tfrac{1}{2}\log\frac{ab}{ab-h^2} = \tfrac{1}{2}\log\frac{1}{1-\left(\frac{h}{\sqrt{ab}}\right)^2} = \tfrac{1}{2}\log\frac{1}{1-r^2} .$$

Thus: $$r = \sqrt{1-e^{-2r_0}} .$$

IV) Correlation coefficient and correlation ratio.

Let X, Y be r.v.'s with finite variances; define correlation coefficient $R(X,Y) = \frac{E(XY)-E(X)E(Y)}{D(X)D(Y)}$, where $D(X)$, $D(Y)$ are the standard deviations of X, Y, i.e. positive roots of $Var(X)$ and $Var(Y)$. Assume $Var(X), Var(Y) > 0$, i.e. X and Y are non-degenerate. By the Cauchy-Schwartz inequality $[E(X - E(X))(Y - E(Y))]^2 \le E(X-EX)^2E(Y-EY)^2$, $|R(X,Y)| \le 1$; $R(X,Y) = 0 \Rightarrow Cov(X,Y) = 0$.

In the latter case we say that X and Y are uncorrelated; this implies that X and Y are independent if (X,Y) is bivariate normal; $R(X,Y) = \pm 1$ if X and Y are directly proportional. Kolmogorov (see Renyi, 1959) defines *correlation ratio* — as a measure of dependence — as follows:

$$K_X(Y) = \frac{D(E(Y|X))}{D(Y)}.$$

The relationship between $R(X,Y)$ and $K_X(Y)$ is given by

<u>Theorem 2</u>: If X and Y are random variables and $\mathrm{Var}(Y)$ exists, then: $K_X(Y) = \sup_g |R(Y, g(X))|$, where g runs over all Borel-measurable real-valued functions $y = g(x)$ such that the variance of $g(X)$ exists. Moreover, $K_X(Y) = |R(Y, g(X))|$ iff $g(X) = a\,E(Y|X) + b$ (a.s.), where $a \neq 0$, b are constants.

(Remark: Notice that the theorem implies that $0 \leq K_X(Y) \leq 1$.)

<u>Proof</u>: Observe that $R(X,Y)$ and $K_X(Y)$ are both invariant under linear transformations. We thus may assume that $E(Y) = 0$, $E(g(X)) = 0$ and $D(Y) = D(g(X)) = 1$. Now $R(Y, g(X)) = E(g(X)Y)$ and since

$$E(g(X)Y|X) = g(X)E(Y|X),$$
$$E(g(X)Y) = E\Big[E(g(X)Y|X)\Big] = E(g(X)E(Y|X)),$$

we have

$$R^2(Y, g(X)) = E^2(g(X)E(Y|X)) \leq \quad \text{(by the C.S. inequality)}$$
$$\leq E(g^2(X))E[E^2(Y|X)] = E[E^2(Y|X)].$$

Hence $R^2(Y, g(X)) \leq E(E^2(Y|X))$.

Since $0 = E(Y) = E(E(Y|X))$, we obtain:

$$R^2(Y, g(X)) \leq E(E^2(Y|X)) - E^2(E(Y|X)) = \mathrm{Var}(E(Y|X))$$

$$\therefore \quad R(Y, g(X)) \leq D(E(Y|X)), \quad \text{or}$$

$$\sup_g |R(Y, g(X))| \le D(E(Y|X)).$$

Now let g_0 be a real function, such that $g_0(X) = E(Y|X)$. (Recall that $E(Y) = 0$, $D(Y) = 1$.) In this case

$$R(Y, g_0(X)) = \frac{E(Yg_0(X)) - E(Y)E(g_0(X))}{D(Y)D(g_0(X))} =$$

$$= \frac{E\left[E(Yg_0(X)|X)\right]}{D(E(Y|X))} = \frac{E(g_0(X)E(Y|X))}{D(E(Y|X))} = \frac{E(E^2(Y|X))}{D(E(Y|X))} =$$

$$= \frac{\mathrm{Var}(E(Y|X))}{DE(Y|X)} = D(E(Y|X)).$$

Thus $D(E(Y|X)) = \sup |R(Y, g(X))|$, where g runs over all real functions such that $\mathrm{Var}(g(X))$ exists. Since

$$|R(Y, g(X))| = |E(Yg(X))| = \left|E\left[E(Yg(X)|X)\right]\right| =$$

$$= |E(g(X)E(Y|X))| \le [Eg^2(X)E(E^2(Y|X))]^{\frac{1}{2}}$$

$$= (EE^2(Y|X))^{\frac{1}{2}} = D(E(Y|X)),$$

the equality

$$K_X(Y) = D(E(Y|X)) = |R(Y, g(X))| \quad \text{holds iff}$$

$$g(X) = aE(Y|X) + b \quad \text{(a.s.)} \quad \text{for some } a \ne 0 \text{ and } b.$$

V) Maximal correlation.

We have $K_X(Y) = \sup_g |R(Y, g(X))|$ if $K_X(Y) = 0$, then $R(Y, g(X)) = 0$ for all g such that $E(g(X))^2 < \infty$. This implies that Y and $g(X)$ are uncorrelated, but does not yet assure that Y and X are independent.

However,

$$S(X,Y) = \sup_{\substack{f,g \\ EF^2(X)<\infty \\ Eg^2(Y)<\infty}} R(f(X), g(Y)) \quad \text{implies}$$

that X and Y *are* independent if $S(X,Y) = 0$.

<u>**Remark**</u>: $S(X,Y)$ also equals $\sup\limits_{\substack{f,g \\ EF^2(X)=Eg^2(Y)=1 \\ Ef(X)=Eg(Y)=0}} R\ f(X),\ g(Y)$ by the fact

that $R(\cdot,\cdot)$ is invariant under linear transformations.

<u>Theorem 3</u>: S_1) $0 \le S(X,Y) \le 1$.

S_2) $S(X,Y) = S(Y,X)$.

S_3) If $\alpha(x)$ and $\beta(y)$ are strictly monotonic, then $S(X,Y) = S(\alpha(X), \beta(Y))$.

S_4) $|R(X,Y)| \le \min(K_X(Y), K_Y(X)) \le \max(K_X(Y), K_Y(X)) \le S(X,Y)$.

S_5) $S(X,Y) = 0$ iff X and Y are independent.

S_6) If there exists an *arbitrary* functional dependence between X and Y, i.e. if there exists Borel measurable functions $f_0(X)$ and $g_0(Y)$ such that $f_0(X)$ is not constant with probability 1 *and* $f_0(X) = g_0(Y)$ then $S(X,Y) = 1$.

<u>Proof</u>: S_1): $0 \le S(X,Y) \le 1$ since $-1 \le R|f(X), g(Y)| \le 1$. The non-negativity of $\sup_{f,g} R(f(X), g(Y))$ comes from the fact that if $Ef(X)g(Y) - Ef(X)Eg(Y) < 0$, we then consider $f' = -f$ to yield $R(f'(X), g(Y)) > 0$.

S_2): $S(X,Y) = S(Y,X)$ in view of the symmetry of $R(X,Y)$.

S_3): Note that in general $Ef^2(X) < \infty \not\Rightarrow Ef^2(\alpha(X)) < \infty$, i.e. $R(\alpha(X), \beta(Y))$ may not exist.

However if α and β are strictly monotonic $\Rightarrow \alpha^{-1}$ and β^{-1} exist and

$$R(f(X), g(Y)) = R(f\alpha^{-1}(\alpha X), g\beta^{-1}(\beta Y))$$

Thus $$\sup_{f,g} R(f(X), g(Y)) \le \sup_{f',g'} R(f'(\alpha X), g'(\beta Y)).$$

Also $$\sup_{f',g'} R(f'(\alpha X), g'(\beta Y)) \le \sup_{f,g} R(f(X), g(Y)).$$

Thus $$S(X,Y) = S(\alpha X, \beta Y).$$

S_4) Since $K_X(Y) = \sup_g |R(Y, g(X))|$, $K_Y(X) = \sup_f |R(f(Y), X)|$,
we have

$$K_X(Y) \geq |R(Y,X)|, \quad K_Y(X) \geq |R(Y,X)|.$$

Hence, $|R(Y,X)| \leq \min(K_X(Y), K_Y(X)) \leq \max(K_X(Y), K_Y(X))$.

$$S(X,Y) = \sup_{f,g} R(f(X), g(Y)) \geq \sup_g R(X, g(Y))$$

and $S(X,Y) \geq \sup_f R(f(X), Y)$.

Thus $|R(X,Y)| \leq \min(K_X(Y), K_Y(X)) \leq \max(K_X(Y), K_Y(X)) \leq S(X,Y)$.

S_5) Define indicator functions on R_1: $f_A(x) = 1 \quad x \in A$; $= 0$ otherwise

$g_B(y) = 1 \quad y \in B$; $= 0$ otherwise, where A and B are arbitrary Borel sets on the real line such that $0 < P(X\in A) < 1$ and $0 < P(Y \in B) < 1$.

By definition:

$$R(f_A(X), g_B(Y)) = \frac{P(X\in A, Y\in B) - P(X\in A)P(Y\in B)}{\sqrt{P(X\in A)(1-P(X\in A))P(Y\in B)(1-P(Y\in B))}}.$$

Hence $S(X,Y) = 0 \Rightarrow P(X\in A, Y\in B) = P(X\in A)P(Y\in B) \Rightarrow X$ and Y are independent.

S_6) If $f_0(X) = g_0(Y)$, consider

$$f_1(X) = \frac{f_0(X)}{1+|f_0(X)|} = \frac{g_0(Y)}{1+|g_0(Y)|} = g_1(Y).$$

(Note that $f_1(X)$ ($= g_1(X)$) are bounded, the variances exist and are $\neq 0$.)

We have $R(f_1(X), g_1(Y)) = \dfrac{Ef_1(X)g_1(Y) - Ef_1(X)Eg_1(Y)}{\sqrt{\operatorname{Var} f_1(X) \operatorname{Var} g_1(Y)}} = 1$, i.e. $S(X,Y) = 1$.

Remark: S_6 is a *sufficient* condition for $S(X,Y) = 1$.

VI) Mean-square contingency.

Let X, Y be arbitrary r.v.'s on (Ω,F,P). The distribution of (X,Y) denoted by P_{XY} is defined on the plane by $P_{X,Y}(C) = P\big((X,Y) \in C\big)$, $C \in \mathcal{B}_2$; $\mathcal{B}_2$ being a two-dimensional Borel σ-algebra. When $C = (-\infty,x] \times (-\infty,y]$ we have the "usual" distribution function:

$$F_{XY}(x,y) = P_{XY}\big((-\infty,x] \times (-\infty,y]\big) = P(X \le x,\ Y \le y).$$

If P_{XY} is absolutely continuous w.r. to $PX^{-1} \times PY^{-1}$, the product measure induced by (X,Y), we have according to the Radon-Nikodym theorem,

$$P_{XY}(C) = \int_C K(x,y)\,dPX^{-1}dPY^{-1}\ , \tag{1}$$

where $K(x,y)$ a Borel measurable function. The R-N derivative $K(x,y)$ is expressed symbolically as

$$dP_{XY}/dPX^{-1}dPY^{-1}\ .$$

We now define

$$\phi(X,Y) = \left[\int_{R^2} \left(\frac{dP_{XY}}{dPX^{-1}dPY^{-1}} - 1\right)^2 dPX^{-1}dPY^{-1}\right]^{1/2} \tag{2}$$

as the *mean square contingency* of X, Y.

"Historical" motivation for this definition:

Let X and Y be discrete r.v.'s with $P(A_K) = P(X=K)$, $K = 1,\ldots S$, $P(B_j) = P(Y=j)$, $j = 1,\ldots,r$. Then:

$$\phi(X,Y) = \left[\sum_i \sum_j \frac{\left(P(A_iB_j) - P(A_i)P(B_j)\right)^2}{P(A_i)P(B_j)} \right]^{\frac{1}{2}} .$$

If we have the $r \times s$ contingency table:

	1	2	3	$\cdots\cdot$	s	
1	ν_{11}	ν_{12}			ν_{1s}	$\nu_{1\cdot}$
2	ν_{21}	ν_{22}			ν_{2s}	$\nu_{2\cdot}$
3	$\cdot$	$\cdot$				
r	ν_{r1}	ν_{r2}	$\cdot$	$\cdots\cdot$	ν_{rs}	$\nu_{r\cdot}$
	$\nu_{\cdot 1}$	$\nu_{\cdot 2}$			$\nu_{\cdot s}$	n

and we estimate $P(A_iB_j)$, $P(A_i)$, $P(B_j)$ by $\frac{\nu_{ij}}{n}$, $\frac{\nu_{i\cdot}}{n}$, $\frac{\nu_{\cdot j}}{n}$, resp. then the estimated ϕ^2 will be:

$$\sum_{i,j} \frac{\left(\nu_{ij}/_n - \nu_{i\cdot}/_n \cdot \nu_{\cdot j}/_n\right)^2}{\nu_{i\cdot}/_n \cdot \nu_{\cdot j}/_n} .$$

This is the statistic for the χ^2 test of independence (ϕ^2 is asymptotically χ^2 distributed with $(s-1) \times (r-1)$ degrees of freedom). Returning to (2), we have

$$\phi(X,Y) = 0 \Rightarrow \frac{dP_{XY}}{dPX^{-1}dPY^{-1}} - 1 = 0 \quad \text{a.s.}$$

with respect to $PX^{-1} \times PY^{-1}$. In other words $K(x,y) = 1$ a.s. w.r. to $PX^{-1} \times PY^{-1}$. From (1) we have in this case

$$P_{XY}(C) = \int_C 1 \, dPX^{-1}dPY^{-1} = PX^{-1} \times PY^{-1}(C) \quad \text{for all} \quad C \in \mathcal{B}_2 .$$

This implies that X, Y are independent. We thus have

Theorem: $\phi(X,Y) = 0 \Longleftrightarrow X, Y$ are independent.

Definition: If X, Y are such that the measure $P(X,Y)^{-1}$ is absolutely continuous w.r. to $PX^{-1} \times PY^{-1}$ (i.e. the mean square contingency exists), it is said that X, Y are *regularly dependent*.

Theorem: If X, Y are regularly dependent, then

$$S(X,Y) \le \phi(X,Y).$$

Proof: Since $0 \le S(X,Y) \le 1$ we can assume that $\phi(X,Y) < 1$. Take measurable f and g such that

$$Ef(X) = Eg(Y) = 0 \quad \text{and} \quad \operatorname{Var} f(X) = \operatorname{Var} g(Y) = 1.$$

Direct calculations show • that

• Indeed

$$R(f(X), g(Y)) = \int_{R^2} f(x)g(y)dP(X,Y)^{-1} =$$

$$= \int_{R^2} f(x)g(y)d\left(P(X,Y)^{-1} - PX^{-1}PY^{-1}\right) =$$

$$= \int_{R^2} f(x)g(x)\left(\frac{dP(X,Y)^{-1}}{dPX^{-1}dPY^{-1}} - 1\right)dPX^{-1}dPY^{-1} .$$

(Observe that both $\displaystyle\int_C \frac{d\left(P(X,Y)-PX^{-1}PY^{-1}\right)}{dPX^{-1}dPY^{-1}} dPX^{-1}dPY^{-1}$

and $\displaystyle\int_C \left(\frac{dP(X,Y)^{-1}}{dPX^{-1}dPY^{-1}} - 1\right)dPX^{-1}dPY^{-1} = P\left((X,Y)\epsilon C\right) -$

$P(X\epsilon C)P(Y\epsilon C)$, for $C\epsilon B_2$)

Hence $\displaystyle R(f(X), g(Y))^2 \le \int_{R^2} f^2(x)g^2(y)dPX^{-1}dPY^{-1} \int_{R^2} \left(\frac{dP(X,Y)^{-1}}{dPX^{-1}dPY^{-1}} - 1\right)^2 dPX^{-1}dPY^{-1}$

$$= \phi^2(X,Y) \quad \text{(by C.-S. inequality).}$$

$$R\big(f(X),\ g(Y)\big)^2 \le \phi^2(X,Y)$$

Hence
$$\sup_{f,g} \left|R\big(f(X),\ g(Y)\big)\right| \le \phi(X,Y)$$

or
$$S(X,Y) \le \phi(X,Y).$$

VII) Pairwise loosely dependent random variables.

<u>Definition</u>: We say that $\{X_n\}$ is a sequence of *pairwise loosely dependent* r.s.'s with coefficient $C>0$ if

$$\left|\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} S(x_n,\ x_m) u_m u_n\right| \le C \sum_{i=1}^{\infty} u_i^2 \qquad \bullet$$

for each sequence $\{u_n\}$ such that $\sum_{i=1}^{\infty} u_i^2 < \infty$.

<u>Theorem</u>: $\{X_n\}$ is an independent family of r.v.'s iff $C=1$.

<u>Proof</u>: $S(X_n,\ X_m) = \begin{cases} 1 & n=m \\ 0 & n\neq m \end{cases}$ if $\{X_n\}$ is independent. We thus can take $C=1$ in $\bullet$.

Suppose $C=1$, define $u_n' = |u_n|$ and assume $\sum u_n^2 < \infty$.

We have:

$$\left|\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} S(X_n,\ X_m) u_n' u_m'\right| \le \sum_{n=1}^{\infty} u_n'^2 \left(= \sum_{n=1}^{\infty} u_n^2\right).$$

Since $0 \le S(X_n,\ X_m)$ and $S(X_n,\ X_n) = 1$, and

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} S(X_n, X_m)u'_n u'_m \le \sum_{n=1}^{\infty} u_n'^2$$

we have $\sum_{n=1}^{\infty} u_n'^2 + \sum_{n\neq m} S(X_n, X_m)u'_n u'_m \le \sum_{n=1}^{\infty} u_n'^2 \Rightarrow S(X_n, X_m) = 0,\ n\neq m \Rightarrow \{X_n\}$ is an independent family.

The *coefficient of dependence* C has another property as well.

<u>Theorem</u>: Let $\{X_n\}$ be a sequence of loosely dependent r.v.'s with coefficient of dependence C. If Y is an arbitrary random variable such that EY^2 exists. Then

$$\sum_{n=1}^{\infty} K_{X_n}^2 \le C.$$

For the proof of the theorem the following lemma is required:

<u>Lemma</u>: Let $\{X_n\}$ be a sequence of square integrable r.v.'s with bound $\theta>0$, i.e.

$$\left|\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} E(X_n X_m)u_n u_m\right| \le \theta \sum_{n=1}^{\infty} u_n^2$$

for every sequence $\{u_n\}$ such that $\sum u_n^2 < \infty$. Then we have for any random variable Y with $EY^2 < \infty$,

$$\sum_{n=1}^{\infty} E^2(YX_n) \le \theta\, EY^2 .$$

<u>Proof</u>: Consider

$$E\Big(Y - \frac{1}{\theta}\sum_{n=1}^{\infty} E(YX_n)X_n\Big)^2 \ge 0,$$

then $E\left\{Y^2 - \frac{2}{\theta}\sum_{n=1}^{\infty} X_n\, Y\, E(YX_n) + \frac{1}{\theta^2}\Big(\sum_{n=1}^{\infty} X_n\, E(YX_n)\Big)^2\right\} \ge 0$

$$\Rightarrow EY^2 \ge \frac{2}{\theta}\sum_{n=1}^{\infty} E^2(YX_n) - \frac{1}{\theta^2} E\left(\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} E(YX_n)X_n X_m E(YX_m)\right)$$

$$\Rightarrow EY^2 \geq \frac{2}{\theta}\sum_{n=1}^{\infty} E^2(YX_n) - \frac{1}{\theta^2}\sum_{n,m} E(YX_n)E(YX_m)E(X_nX_m)$$

$$\geq \frac{2}{\theta}\sum_{n=1}^{\infty} E^2(YX_n) - \frac{\theta}{\theta^2}\sum_{n=1}^{\infty} E(YX_n)^2 = \frac{1}{\theta}\sum_{n=1}^{\infty} E^2(YX_n).$$

(Remark: In Hilbert spaces theory, for any orthonormal system of r.v.'s $\{X_n\}$, i.e. $EX_nX_m = 0$, $n \neq m$ $(EX_n^2)^{1/2} = 1$, we have $\theta = 1$ and $EY^2 \geq \sum_{n=1}^{\infty} E^2YX_n$ (*the so-called Bessel inequality*). The last lemma can be viewed as a generalization of the Bessel inequality.)

To *prove* the theorem observe that

$$K_{X_n}(Y) \equiv \frac{D(E(Y|X_n))}{D(Y)} = \sup_{\substack{g \\ Eg^2(X_n)<\infty}} R(Y, g(X_n)).$$

Let $\{g_n\}$ be a sequence of measurable functions such that $Eg_n^2(X_n) < \infty$. (Such a sequence can always be constructed by choosing $g_n(X_n) = \frac{|f(X_n)|}{1+|f(X_n)|}$, for any measurable f.) Define $g_n'(X_n) = \frac{g_n(X_n)-E(g_n(X_n))}{D(g_n(X_n))}$, $n = 1,2\ldots$.

Since $\{X_n\}$ is loosely dependent with coefficient of dependence C we have

$$\left|\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} E(g_n'(X_n)g_m'(X_m)u_nu_m\right| = \left|\sum_n\sum_m R(g_n(X_n)g_m(X_m))u_nu_m\right|$$

$$\leq \sum_n\sum_m S(X_n,X_m)|u_n||u_m| \leq C\sum_{n=1}^{\infty} u_n^2 .$$

Using the Lemma with $Y' = Y - EY$, we have:

$$\sum_{n=1}^{\infty} E^2Y'g_n'(X_n) \leq CEY'^2 = CD^2(Y).$$

If we choose $g_n(X_n)$ in a special way to be equal to

$$g_n(X_n) = E(Y'|X_n),$$

we obtain

$$\sum_{n=1}^{\infty} E^2\left[Y' \frac{E(Y'|X_n)-EY'}{D(E(Y'|X_n))}\right] \le CD^2(Y);$$

or

$$\sum_{n=1}^{\infty} \frac{E^2[Y'E(Y'|X_n)]}{D^2(E(Y'|X_n))} \le CD^2(Y). \qquad \bullet$$

Note that:

$$E(Y'E(Y'|X_n)) = E\Big\{E(Y'E(Y'|X_n)|X_n)\Big\}$$

$$= E(E(Y'|X_n)E(Y'|X_n)) = E(E^2(Y'|X_n)) = D^2(E(Y'|X_n)),$$

since $E^2E(Y'|X_n) = E^2Y' = 0$.

Therefore the l.h.s. of $\bullet$ becomes:

$$\sum_{n=1}^{\infty} D^2(E(Y'|X_n)) \le CD^2(Y). \qquad \bullet\bullet$$

However,

$$E(Y'|X_n) = E(Y - EY|X_n) = E(Y|X_n) - EY.$$

Thus

$$D^2(E(Y'|X_n)) = D^2(E(Y|X_n)).$$

We finally obtain from $\bullet\bullet$:

$$\sum_{n=1}^{\infty} D^2(E(Y|X_n)) \le CD^2(Y)$$

or

$$\sum_{n=1}^{\infty} \frac{D^2(E(Y|X_n))}{D^2(Y)} = \sum_{n=1}^{\infty} K^2_{X_n}(Y) \le C. \qquad \square$$

We know that C can be chosen to be 1 provided $\{X_n\}$ is a sequence of pairwise independent r.v.'s.

Corollary: If $\{X_n\}$ is a sequence of pairwise independent r.v.'s, and Y is an arbitrary r.v. such that $EY^2 < \infty$, then

$$\sum_{n=1}^{\infty} K^2_{X_n}(Y) \le 1.$$

VIII) Positive dependence and multivariate hazard rates.

Given an n-dimensional random variable $\underset{\sim}{X} = (X_1,\ldots,X_m)$, denote by

$$F_{\underset{\sim}{X}}(\underset{\sim}{x}) = \Pr[\bigcap_{j=1}^{m} (X_j \le x_j)] \text{ the joint c.d.f. and by}$$

$$G_{\underset{\sim}{X}}(\underset{\sim}{x}) = \Pr[\bigcap_{j=1}^{m} (X_j > x_j)] \text{ the joint survival function.}$$

We assume that $F_{\underset{\sim}{X}}(\underset{\sim}{x})$ and $G_{\underset{\sim}{X}}(\underset{\sim}{x})$ are absolutely continuous.

Definition.

If $h_{\underset{\sim}{X}}(\underset{\sim}{x})_j$ is an increasing (decreasing) function of x_j, for $j = 1,2,\ldots,m$, for all $\underset{\sim}{x} \in R^m$, then the distribution $F_{\underset{\sim}{X}}(\underset{\sim}{x})$ is a (vector)-multivariate *increasing* hazard rate [IHR] (decreasing hazard rate [DHR]) distribution, where $h_{\underset{\sim}{X}}(\underset{\sim}{x})_j$ is the j-th component of -grad log $G_{\underset{\sim}{X}}(\underset{\sim}{x})$.

Lemma 1: If $X_1,\ldots X_m$ are mutually independent i.e.

$$G_{\underset{\sim}{X}}(\underset{\sim}{x}) = \prod_{j=1}^{m} G_{X_j}(x_j)$$

then $h_X(x)_j = h_{X_j}(x_j)$, where $h_{X_j}(x_j) \equiv -\frac{d}{dx_j} \log G_{X_j}(x_j) = f_{X_j}(x_j)/1-F_{X_j}(x_j)$ is the *univariate* hazard of the j-th component.

Lemma 2: If $h_{\underset{\sim}{X}}(\underset{\sim}{x}) = \underset{\sim}{c}$ where $\underset{\sim}{c} = (c_1,\ldots,c_m)$ is an absolute constant, then $\underset{\sim}{X} = (X_1,\ldots X_n)$ are mutually independent exponential random variables

and conversely.

Proof: Given $h_{\underset{\sim}{X}}(\underset{\sim}{x}) = \underset{\sim}{c}$, this implies that:

$$\frac{\partial \log G_{\underset{\sim}{X}}(\underset{\sim}{x})}{\partial x_j} = -c_j \qquad (j = 1,\ldots m)$$

i.e. $G_{\underset{\sim}{X}}(\underset{\sim}{x}) = e^{-c_j x_j} g_j(x_1,\ldots x_{j-1}, x_{j+1},\ldots x_m), \quad \forall\, j = 1,\ldots,m,$

i.e. $$G_{\underset{\sim}{X}}(\underset{\sim}{x}) \propto \exp\Big(-\sum_j c_j x_j\Big).$$

The boundary conditions on $G_{\underset{\sim}{X}}(\underset{\sim}{x})$ imply that

$$G_{\underset{\sim}{X}}(\underset{\sim}{x}) = \exp\Big(-\sum_j^m c_j x_j\Big), \qquad x_j \geq 0. \qquad \square$$

Recall from Section II that the variables $\underset{\sim}{X} = (X_1, X_2, \ldots, X_m)$ are G-*positively* quadrant dependent if

$$R_G(\underset{\sim}{x}) \equiv G_{\underset{\sim}{X}}(\underset{\sim}{x}) / \prod_{i=1}^{m} G_{X_i}(x_i) \geq 1.$$

Observe that $\lim\limits_{x_1 \to -\infty} R_G(x_1,\ldots,x_m) = R_G(x_2,\ldots,x_m)$. In particular $\lim\limits_{x_1 \to -\infty} R_G(x_1,x_2) = 1$ (for $m=2$). Moreover $\dfrac{\partial \log R_G(\underset{\sim}{x})}{\partial x_j} = h_{X_j}(x_j) - h_{\underset{\sim}{X}}(\underset{\sim}{x})_j$, $j = 1,\ldots m$.

Theorem: For $m=2$, if $h_{X_j}(x_j) > h_{\underset{\sim}{X}}(\underset{\sim}{x})_j$ $(j = 1,2)$ for all $\underset{\sim}{x}$, then $R_G(\underset{\sim}{x}) \geq 1$ for all $\underset{\sim}{x}$, i.e. the variables X_1 and X_2 are G-*positively* quadrant dependent

Proof: Observe that $\dfrac{\partial \log R_G(\underset{\sim}{x})}{\partial x_j} > 0$ for all $\underset{\sim}{x}$ and $j = 1,2$ together with $\lim\limits_{x_1 \to -\infty} R_G(x) = 1$ and $\lim\limits_{x_2 \to -\infty} R_G(x) = 1$, implies that $R_G(\underset{\sim}{x}) \geq 1$ for all $\underset{\sim}{x}$. $\square$

The converse of this theorem is in general *not true*.

Consider the survival function:

$$G_{12}(x_1,x_2) = G_1(x_1)G_2(x_2)[1 + \alpha e^{-\frac{1}{2}(x_1^2+x_2^2)}] \quad \underset{\substack{-\infty<x_1<\infty \\ -\infty<x_2<\infty}}{(\alpha>0)} .$$

It is a survival function for some choices of $G_1(x_1)$ and $G_2(x_2)$ and certain values of α. Indeed, first we have to assure that

$\frac{\partial^2 F_{12}}{\partial x_1 \partial x_2} = \frac{\partial^2 G_{12}}{\partial x_1 \partial x_2} > 0$ for all x_1 and x_2 .

$$\text{Now} \quad \frac{\partial^2 F_{12}}{\partial x_1 \partial x_2} = \frac{\partial^2 G_{12}}{\partial x_1 \partial x_2} = f_1(x_1)f_2(x_2) + \alpha e^{-\frac{1}{2}(x_1^2+x_2^2)} \times$$
$$\times [f_1(x_1) + x_1G_1(x_1)][f_2(x_2) + x_2G_2(x_2)]$$
$$\geq f_1(x_1)f_2(x_2) - \alpha|x_1x_2|e^{-\frac{1}{2}(x_1^2+x_2^2)} ,$$

where $\frac{\partial^2 F_{12}}{\partial x_i} = f_i$, $i = 1,2$. Taking $f_j(x_j) = \frac{1}{2}|x_j|e^{-\frac{1}{2}x_j}$, we have

$\frac{\partial^2 F_{12}}{\partial x_1 \partial x_2} \geq (\frac{1}{4}-\alpha)|x_1x_2|e^{-\frac{1}{2}(x_1^2+x_2^2)} > 0$ if $0 < \alpha < \frac{1}{4}$. Also: $G_{12}(x_1,x_2) \to 0$ as $x_1 \to +\infty$; $\to 0$ as $x_2 \to +\infty$; $\to G_i(x_i)$ as $x_{3-i} \to -\infty$ $_{i=1,2}$. For this survival function we have:

(i) $R_G = \frac{G_{12}}{G_1G_2} = 1 + \alpha e^{-\frac{1}{2}(x_1^2+x_2^2)} > 0$, so we have G-positive quadrant dependence, but

(ii) R_G is an increasing function of x_1 for $x_1 > 0$ and decreasing function of x_1 for $x_1 > 0$, so $h_{\underset{\sim}{X}}(\underset{\sim}{x})_1 - h_{X_1}(x_1)$ changes sign as x_1 increases. However for a family of bivariate survival distributions of the form:

$$G_{X_1,X_2}(x_1,x_2) = G_{X_1}(x_1)G_{X_2}(x_2)\Big[1 + \alpha F_{X_1}(x_1)F_{X_2}(x_2)\Big], \quad |\alpha| \leq 1,$$

(note that the distribution in the counterexample above does *not* belong to this family). We have $R_G > 1 \iff h_{X_j}(x_j) - h_{\underset{\sim}{X}}(\underset{\sim}{x})_j > 0, \quad j = 1,2.$ The proof is presented in Appendix 4.

IX) Measures of dependence via "copulas".

a) Definition and properties of "copulas".

The *quadrant dependence* measures the deviation of the bivariate distribution from its marginals.

A more general problem is to relate (explicitly) a multivariate distribution function to its marginals. The FGM family (discussed at the end of the previous section) expresses the joint bivariate distribution as an explicit function of the marginals:

$$F_{XY}(x,y) = F_X(x)F_Y(y)[1 + \alpha(1 - F_X(x))(1 - F_Y(y))], \qquad |\alpha| \le 1.$$

Other examples of this situation are:

$$F^*_{XY}(x,y) = \max(0, F_X(x) + F_Y(y) - 1) \qquad \text{(the lower } \textit{Frechét bound}\text{)}$$

$$F^{**}_{XY}(x,y) = \min\left[F_X(x, F_Y(y))\right] \qquad \text{(the upper Frechét bound)}$$

(or any linear combination with positive weights adding up to 1 of $F^*_{XY}(x,y)$ and $F^{**}_{XY}(x,y)$) and of course the independent case

$$F^{\circ}_{XY}(x,y) = F_X(x) \cdot F_Y(y).$$

The definition and the theorem below present an answer to the following two questions:

1) Can a given multivariate distribution function be represented as a function of its marginals?
2) What are the characteristics of this function if the answer to 1) is affirmative?

Definition: A *copula* C is a real-valued function of n variables ($n \geq 2$) defined on a subset of $[0,1] \times [0,1] \ldots \times [0,1]$, with the range being a subset of the interval $[0,1]$, satisfying the following properties:

(C_1) $C(1,\ldots,1, x_m, 1,\ldots,1) = x_m$, $m \leq n$, $x_m \in [0,1]$,

(C_2) $C(x_1,x_2,\ldots,x_n) = 0$ if $x_m = 0$ for any $m \leq n$,

(C_3) C is non-decreasing in each variable.

Theorem: For $n \geq 2$, let F be an n-dimensional distribution function with marginals $F_1,F_2,\ldots,F_n$. Then there exists a copula C such that

$$F(x_1,x_2,\ldots x_n) = C(F_1(x_1),\ldots,F_n(x_n))$$

for all n-tuples $(x_1,x_2,\ldots,x_n) \in R^n$.

Proof: To show that F is a function of $F_1,F_2,\ldots,F_n$, consider any two points $\underset{\sim}{x} = (x_1,x_2,\ldots,x_n)$ and $\underset{\sim}{y} = (y_1,\ldots y_n) \in R^n$.

We have

$$|F(x_1,x_2,\ldots,x_n) - F(y_1,y_2,\ldots,y_n)| \leq$$
$$\leq |F_1(x_1) - F_1(y_1)| + |F_2(x_2) - F_2(y_2)| + \ldots + |F_n(x_n) - F_n(y_n)|. \bullet$$

This inequality shows that the set of points

$$[(F_1(x_1),\ F_2(x_2)\ \dots\ F_n(x_n)),\ F(x_1,x_2,\dots,x_n)]|\underset{\sim}{x} \in R^n$$

is a graph of a function C (if each F_j is continuous then C is unique). Inequality • implies that $C(x_1,x_2,\dots,x_n)$ is a jointly continuous function of $x_1,x_2,\dots,x_n$. Utilizing the basic properties of distribution functions it can be easily verified that the function C satisfies the properties (C_1), (C_2) and (C_3). □

b) Copulas and dependence.

Let $F_{XY}(u,v) = C_{XY}(F_X(u),\ F_Y(v))$. From Frechét's bounds we obtain:

$$\max(x+y-1,\ 0) \le C_{XY}(x,y) \le \min(x,y) \qquad x,y \in [0,1].$$

If X and Y are independent, we have

$$C_{XY}(x,y) = xy \qquad x,y \in [0,1].$$

These observations suggest that the volume between the two surfaces $Z = C_{XY}(x,y)$ and $Z = xy$ may serve as a measure of dependence between X and Y. This measure is formally defined (Schweizer and Wolff (1976)) by

$$\sigma(X,Y) = K\int_0^1\int_0^1 |xy - C_{XY}(x,y)|\,dxdy,$$

where K is chosen in such a manner that $\sigma(X,Y) \le 1$ for all X and Y. (Observe that $\sigma(X,Y) = 0 \iff$ X and Y are independent. Also $\sigma(X,Y) = \sigma(Y,X)$.)

Direct computations yield:

$$\int_0^1\int_0^1 |xy - 0|\,dxdy = \tfrac{1}{4};$$

$$\int_0^1 \int_0^1 |xy - \min(x,y)|dxdy = \frac{1}{12} .$$

The normalizing constant K therefore equals 12 and $\sigma(X,Y) = 1$ for the Frechet's upper bound $\min(F_X(x), F_Y(y))$. Thus, finally,

$$\sigma(X,Y) = 12 \int_0^1 \int_0^1 |xy - C_{XY}(x,y)|dxdy.$$

According to this measure the *maximal* dependence between the variables is when $F_{X,Y}(x,y) = \min(F_X(x), F_Y(y))$, i.e. when the joint distribution is given by a diagonal-type surface over the "$F_X(x)$, $F_Y(y)$"-plane.

<u>Part 2 Positive dependence revisited</u>.

<u>I) Introduction</u>.

Recall the corollary of Theorem 1 Section II of Part 1 which states that if $(X,Y) \in F_1$ and EXY, EX and EY exist, then $\mathrm{Cov}(X,Y) \geq 0$.

Esary, Proschan and Walkup (1967) define association of X and Y by requiring that $\mathrm{Cov}(f(X,Y), g(X,Y)) \leq 0$ for all non-decreasing real-valued functions f and g. They also present a multivariate version of this definition:

$$\mathrm{Cov}(f(\underset{\sim}{X}), g(\underset{\sim}{X})) \geq 0 \text{ for non-decreasing real-valued } f \text{ and } g \text{ where } \underset{\sim}{X} = (X_1,X_2,\ldots,X_n).$$

In numerous reliability situations, the random variables of interest are usually not independent, but often satisfy the association property.

(*See Barlow and Proschan (1975)*.)

Regression dependence, likelihood ratio dependence were discussed by Lehmann (1966). The TP_2 property of Karlin (1968) is analogous to Lehmann's likelihood ratio dependence, which is useful for determination of the hazard rate behavior in univariate models.

II) Association of random variables.

Definition 1: Random variables $(X_1, X_2, \ldots, X_n)$ are called *associated* if $\mathrm{Cov}\big(f(\underset{\sim}{X}), g(\underset{\sim}{X})\big) \geq 0$ for all non-decreasing functions f and g such that $Ef(\underset{\sim}{X})$, $Eg(\underset{\sim}{X})$, and $Ef(\underset{\sim}{X})g(\underset{\sim}{X})$ exist.

Remark: f, g are non-decreasing if they are non-decreasing for each variable when the rest are *held fixed*.

Property 2: Any subset of associated random variables forms a set of associated variables.

Property 3: If two sets of associated r.v.'s are independent of one another, then their union is a set of associated r.v.'s.

Proof: Let $\underset{\sim}{X} = (X_1, \ldots, X_n)$ and $\underset{\sim}{Y} = (Y_1, \ldots, Y_m)$ be two sets of associated r.v.'s, let $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ be independent and f and g be non-decreasing functions. We have:

$$\mathrm{Cov}\big(f(\underset{\sim}{X},\underset{\sim}{Y}), g(\underset{\sim}{X},\underset{\sim}{Y})\big) = Ef(\underset{\sim}{X},\underset{\sim}{Y})g(\underset{\sim}{X},\underset{\sim}{Y}) - Ef(\underset{\sim}{X},\underset{\sim}{Y})Eg(\underset{\sim}{X},\underset{\sim}{Y})$$

$$= \int_{R^m}\int_{R^n} f(\underset{\sim}{x},\underset{\sim}{y})g(\underset{\sim}{x},\underset{\sim}{y})\,dp\underset{\sim}{X}^{-1}dp\underset{\sim}{Y}^{-1}$$

$$- \int_{R^m}\int_{R^n} f(\underset{\sim}{x},\underset{\sim}{y})\,dp\underset{\sim}{X}^{-1}dp\underset{\sim}{Y}^{-1} \,.$$

$$\cdot \int_{R^m}\int_{R^n} g(\underset{\sim}{x},\underset{\sim}{y})dpX^{-1}dpY^{-1} \cdot$$

$$= \int_{R^m}\left[\int_{R^n} f(\underset{\sim}{x},\underset{\sim}{y})g(\underset{\sim}{x},\underset{\sim}{y})dp\underset{\sim}{X}^{-1} - \int_{R^n} f(\underset{\sim}{x},\underset{\sim}{y})dp\underset{\sim}{X}^{-1}\int_{R^n} g(\underset{\sim}{x},\underset{\sim}{y})dp\underset{\sim}{X}^{-1}\right]dpY^{-1}$$

$$+ \int_{R^m}\left[\int_{R^n} f(\underset{\sim}{x},\underset{\sim}{y})dpX^{-1}\int_{R^n} g(\underset{\sim}{x},\underset{\sim}{y})dp\underset{\sim}{X}^{-1}\right]dp\underset{\sim}{Y}^{-1}$$

$$- \int_{R^m}\int_{R^n} f(\underset{\sim}{x},\underset{\sim}{y})dp\underset{\sim}{X}^{-1}dpY^{-1}\int_{R^m}\int_{R^n} g(\underset{\sim}{x},\underset{\sim}{y})dp\underset{\sim}{X}^{-1}dp\underset{\sim}{Y}^{-1} =$$

$$= \mathrm{I} + \mathrm{II} - \mathrm{III}.$$

Now $\mathrm{I}\geq 0$ since $X_1,\ldots,X_n$ are associated; also, $\mathrm{II} - \mathrm{III} =$ $\mathrm{Cov}\left(\int_{R^n} f(\underset{\sim}{x},\underset{\sim}{Y})dp\underset{\sim}{X}^{-1}, \int_{R^n} g(\underset{\sim}{x},\underset{\sim}{Y})dp\underset{\sim}{X}^{-1}\right) \geq 0$, because $\int_{R^n} f(\underset{\sim}{x},\underset{\sim}{y})dP\underset{\sim}{X}^{-1}$, $\int_{R^n} g(\underset{\sim}{x},\underset{\sim}{y})dP\underset{\sim}{X}^{-1}$ are non-decreasing in $y_1,y_2,\ldots,y_m$, and the variables $\underset{\sim}{Y} = (Y_1,\ldots,Y_m)$ are associated. □

<u>Property 4</u>: *A set consisting of a single variable is associated.*

<u>Proof</u>: It is required to show that $\mathrm{Cov}\left(f(X), g(X)\right) \geq 0\ \forall$ non-decreasing f and g.

Recall (Section II in Part 1) that Hoeffding's lemma yields

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(G_{XY}(u,v) - G_X(u)G_Y(v)\right)dudv$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{Cov}\left(I(u,X), I(v,Y)\right)dudv,$$

where $I(u,x) = 1$ if $u\leq x$ and 0 otherwise

$I(v,x) = 1$ if $v\leq x$ and 0 otherwise,

and $G_{XY}(u,v) = P(X>u, Y>v)$, $G_X(u) = P(X>u)$,

$$G_Y(v) = P(Y>v).$$

Define now

$$I\big(u,\ f(x)\big) = 1 \quad \text{if} \quad f(x) \ge u \quad \text{and} \quad 0 \quad \text{otherwise}$$

$$I\big(u,\ g(x)\big) = 1 \quad \text{if} \quad g(x) \ge v \quad \text{and} \quad 0 \quad \text{otherwise.}$$

Then as in the proof of Hoeffding's lemma (Lemma 3, Part 1):

$$\mathrm{Cov}\big(f(X),\ g(X)\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{Cov}\Big[I\big(u,\ f(X)\big),\ I\big(v,\ g(X)\big)\Big]dudv.$$

Observe that $I\big(u,\ f(x)\big)$ and $I\big(v,\ g(x)\big)$ are two-valued non-decreasing functions of x. There are two possibilities, either $I\big(u,\ f(x)\big) \ge I\big(v,\ g(x)\big)$ for all x, or $I\big(u,\ f(x)\big) < I\big(v,\ g(x)\big)$ for all x.

In the first case:

$$\mathrm{Cov}\big(f(X),\ g(X)\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Big\{E[I\big(u,\ f(X)\big)I\big(v,\ g(X)\big)] - E\Big[I\big(u,\ f(X)\big)\Big] \times$$
$$\times\ E\Big[I\big(v,\ g(X)\big)\Big]\Big\}dudv$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[I\big(u,\ g(X)\big)] - E\Big[I\big(u,\ f(X)\big)\Big]E\Big[I\big(v,\ g(X)\big)\Big]dudv$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\Big[I\big(v,\ g(X)\big)\Big][1 - E\Big[I\big(u,\ f(X)\big)\Big]]dudv \ge 0.$$

In the second case:

$$\mathrm{Cov}\big(f(X),\ g(X)\big) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\Big[I\big(u,\ f(X)\big)\Big]\ 1 - E\Big[I\big(v,\ g(X)\big)\Big]dudv \ge 0$$

and Prop. 4 is verified. □

<u>Property 5</u>: **Non-decreasing functions of associated random variables are associated.**

<u>Proof</u>: Let $X_1, X_2, \ldots, X_n$ be associated random variables; h_i $(i = 1,\ldots,m)$ be non-decreasing functions of n variables. To show that $Y_i = h_i(\underset{\sim}{X})$,

$i = 1,\ldots,m$, are associated it is sufficient to show that $\mathrm{Cov}(f(\underset{\sim}{Y}), g(\underset{\sim}{Y})) \geq 0$ for all non-decreasing f and g of m variables.

However, $\mathrm{Cov}(f(\underset{\sim}{Y}), g(\underset{\sim}{Y})) = \mathrm{cov}\left[f(\underset{\sim}{h}(\underset{\sim}{X})), g(\underset{\sim}{h}(\underset{\sim}{X}))\right]$ and the latter is non-negative since $X_1,\ldots,X_n$ are associated. □

<u>Property 6</u>: Independent random variables are associated.

<u>Proof</u>: Let $X_1,X_2,\ldots,X_n$ be independent random variables.

By Prop. 4 X_1 is associated and X_2 is associated.

By Prop. 3 (X_1,X_2) are associated, and the result follows by induction. □

<u>III) Positive regression dependence</u>.

Recall that $(X,Y) \in F_1$, i.e. X and Y have positive quadrant dependence (notation $\mathrm{PQD}(X,Y)$), if $P(X\leq x, Y\leq y) \geq P(X\leq x)P(Y\leq y)$ for all $x,y \in R_1$. If $P(X\leq x) > 0$, this condition can be restated as $P(Y\leq y|X\leq x) \geq P(Y\leq y)$, $\forall$ x,y. This observation motivates the following two notions:

<u>Definition 2</u>: If $P(Y\leq y|X\leq x) \downarrow$ as $x \uparrow$ for all y, we say that Y is *left tail decreasing* in X (notation $\mathrm{LTD}(Y|X)$),

<u>Definition 3</u>: If $P(Y\leq y|X=x) \downarrow$ as $x \uparrow$ for all y, we say that Y is *positively regression dependent* on X. (Notation $\mathrm{PRD}(Y|X)$ *or* $(X,Y) \in F_2$.)

<u>Example 1</u> (Lehmann 1966): Let $Y = \alpha + \beta X + U$, where X and U are independent r.v.'s. In this case:

$$P(Y\le y|X=x) = E(I_{[Y\le y]}|X=x) = E(I_{[\alpha+\beta X+U\le y]}|X=x)$$

$$= E(I_{[\alpha+\beta x+U\le y]}|X=x) = EI_{[\alpha+\beta x+U\le y]} = P(\alpha + \beta x + U \le y).$$

Thus, $P(Y\le y|X=x) \downarrow$ as $x \uparrow$ if $\beta > 0$ $[PRD(Y|X)]$

$P(Y\le y|X=x) \uparrow$ as $x \uparrow$ if $\beta < 0$ $[NRD(Y|X)]$

and $P(Y\le y|X=x) = P(\alpha + U \le y)$ if $\beta=0$.

Theorem 1: $PRD(Y|X) \Rightarrow LTD(Y|X) \Rightarrow PQD(X,Y)$.

Proof: From Def. 2, $P(Y\le y|X\le x) \ge P(Y\le y|X\le x')$, $x < x'$.

Let $x' \uparrow \infty$, then

$$P(Y\le y|X\le x) \ge \lim_{x'\uparrow\infty} \frac{P(Y\le y, X\le x')}{P(X\le x')} = P(Y\le y).$$

Thus $LTD(Y|X) \Rightarrow PQD(X,Y)$.

Now, if $LTD(Y|X)$ holds then

$$P(Y\le y|X\le x) \ge P(Y\le y|X\le x') \qquad \text{for all } x<x' \text{ and all } y.$$

i.e. $$\frac{\int_{-\infty}^{x} P(Y\le y|X=u)dP(X\le u)}{P(X\le x)} \le \frac{\int_{-\infty}^{x'} P(Y\le y|X=u)dP(X\le u)}{P(X\le x')},$$

with $x < x'$. If $P(Y\le y|X=u)$ is a decreasing function of u (i.e. $pRD(Y|X)$ holds) then the last inequality is valid. Thus $PRD(Y|X) \Rightarrow LTD(Y|X)$. □

IV) Relationships between some notions of bivariate positive (negative) dependence.

In addition to the three definitions of bivariate dependence introduced in Section 3) we define:

Definition 4: Y is *right tail increasing* in X $\left(RTI(Y|X)\right)$ if

$$P(Y>y|X>x) \uparrow \text{ as } x \uparrow \text{ for all } y.$$

Definition 5: Y is *stochastically increasing* in X $\left(SI(Y|X)\right)$ if

$$P(Y>y|X=x) \uparrow \text{ as } x \uparrow \text{ for all } y.$$

Definition 6: If X, Y have joint density $f(x,y)$, we say that $f(x,y)$ is TP$_2$ or TP$_2(X,Y)$ if

$$\begin{vmatrix} f(x_1,y_1) & f(x_1,y_2) \\ f(x_2,y_1) & f(x_2,y_2) \end{vmatrix} \geq 0 \qquad \begin{array}{l} \text{for all } x_1 < x_2 \text{ and } y_1 < y_2 \\ \text{in the domain of } X \text{ and } Y. \end{array}$$

In addition to the relations – due to Lehmann (1966) –

$$PRD \Rightarrow LTD \Rightarrow PQD$$

which was proved in the previous section, the following has been shown by Esary and Proschan.

Theorem 2 (Esary and Proschan 1972):

$$TP_2(X,Y) \Rightarrow SI(Y|X) \Rightarrow RTI(Y|X) \Rightarrow PQD(X,Y).$$

Proof: If $TP_2(X,Y)$ is valid, then by definition

$$\begin{vmatrix} f(x_1,y_1) & f(x_1,y_2) \\ f(x_2,y_1) & f(x_2,y_2) \end{vmatrix} \geq 0 \qquad \begin{matrix} x_1 < x_2 \\ y_1 < y_2 \end{matrix},$$

or $f(x_1,y_1)f(x_2,y_2) \geq f(x_1,y_2)f(x_2,y_1)$.

Integrating over the variables y_1 and y_2 we have:

$$\int_{-\infty}^{y} f(x_1,y_1)dy_1 \int_{y}^{\infty} f(x_2,y_2)dy_2 \geq \int_{-\infty}^{y} f(x_2,y_1)dy_1 \int_{y}^{\infty} f(x_1,y_2)dy_2$$

or

$$\begin{vmatrix} \int_{y}^{\infty} f(x_1,z)dz & \int_{y}^{\infty} f(x_2,z)dz \\ \int_{-\infty}^{y} f(x_1,z)dz & \int_{-\infty}^{y} f(x_2,z)dz \end{vmatrix} \geq 0.$$

Adding the 1-st row of the last determinant to the second we obtain

$$\begin{vmatrix} \int_{y}^{\infty} f(x_1,z)dz & \int_{y}^{\infty} f(x_2,z)dz \\ f_1(x_1) & f_2(x_2) \end{vmatrix} \leq 0$$

$$\Longrightarrow \frac{\int_y^\infty f(x_1,z)dz}{f_1(x_1)} \leq \frac{\int_y^\infty f(x_2,z)dz}{f_2(x_2)}$$

or $P(Y>y|X=x_1) \leq P(Y>y|X=x_2)$, $x_1 < x_2$ i.e. $SI(Y|X)$ is valid. Thus $TP_2(X,Y) \Rightarrow SI(Y|X)$. Now let X and Y satisfy $SI(Y|X)$.

By definition:

$$P(Y>y|X=x_1) \leq P(Y>y|X=x_2) \quad \text{for} \quad x_1 < x_2$$

and all y.

Equivalently,

$$\frac{\int_y^\infty f(x_1,z)dz}{f_1(x_1)} \leq \frac{\int_y^\infty f(x_2,z)dz}{f_1(x_2)}$$

or

$$f_1(x_2)\int_y^\infty f(x_1,z)dz \le f_1(x_1)\int_y^\infty f(x_2,z)dz \le 0.$$

Integrating ● from S_1 to S_2 (with $S_1 < S_2$) over x_1 , from S_2 to ∞ over x_2 , we have:

$$\int_{S_1}^{S_2}\int_y^\infty f(x_1,z)dzdx_1 \int_{S_2}^\infty f_1(x_2)dx_2 \le \int_{S_2}^\infty\int_y^\infty f(x_2,z)dzdx_2 \int_{S_1}^{S_2} f_1(x_1)dx_1$$

(the inequality ● remains valid as long as the ranges of integration satisfy $x_1 \le x_2$)

or

$$\begin{vmatrix} \int_{S_1}^{S_2}\int_y^\infty f(x_1,z)dzdx_1 & \int_{S_2}^\infty\int_y^\infty f(x_2,z)dzdx_2 \\ \int_{S_1}^{S_2} f_1(x_1)dx_1 & \int_{S_2}^\infty f_1(x_2)dx_2 \end{vmatrix} \le 0.$$

Adding the second column to the first we have

$$\begin{vmatrix} \int_{S_1}^\infty\int_y^\infty f(x_1,z)dzdx_1 & \int_{S_2}^\infty\int_y^\infty f(x_2,z)dzdx_2 \\ \int_{S_1}^\infty f_1(x_1)dx_1 & \int_{S_2}^\infty f_1(x_2)dx_2 \end{vmatrix} \le 0.$$

This implies that

$$\frac{\int_{S_1}^\infty\int_y^\infty f(x_1,z)dzdx_1}{\int_{S_1}^\infty f_1(x_1)dx_1} \le \frac{\int_{S_2}^\infty\int_y^\infty f(x_2,z)dzdx_1}{\int_{S_2}^\infty f_1(x_2)dx_2},$$

i.e. $P(Y>y|X>S_1) \le P(Y>y|X>S_2)$, $S_1 < S_2$ which means that $RTI(Y|X)$ is valid. To complete the proof of Theorem 2 we must establish that

$$RTI(Y|X) \Rightarrow PQD(X,Y).$$

It is easy to show that $A(X,Y)$ (*association* between X and Y) implies

PQD(X,Y).

Indeed:

$$A(X,Y) \Rightarrow \mathrm{Cov}\big(f(X),\, g(Y)\big) \geq 0$$

for all non-decreasing f and g. Now let

$$f(u) = 1 \quad \text{if} \quad u > x \quad \text{and} \quad 0 \quad \text{otherwise}$$
$$g(v) = 1 \quad \text{if} \quad v > y \quad \text{and} \quad 0 \quad \text{otherwise.}$$

Since these particular f and g are non-decreasing we have

$$\begin{aligned} 0 &\leq \mathrm{Cov}\big(f(X),\, g(Y)\big) = E\big(f(X)g(Y)\big) - Ef(X)Eg(Y) \\ &= P(X>x,\, Y>y) - P(X>x)P(Y>y) = P(X\leq x,\, Y\leq y) - \\ &\quad P(X\leq x)P(Y\leq y). \end{aligned}$$

In other words PQD(X,Y) is implied.

The missing part involves verification of the implication RTI(Y|X) $\Rightarrow$ A(X,Y). The proof of this proposition is quite long and constitutes the major proof of Esary and Proschan's 1972 paper in the *Ann. of Math. Statist.*

Final remark. If the variables X and Y take on values 0 and 1 only, all the above conditions of dependence are equivalent:

Indeed in this case PQD(X,Y) $\Rightarrow$ TP_2(X,Y).

Consider

$$D = \begin{vmatrix} P(X=0,\ Y=0), & P(X=0,\ Y=1) \\ P(X=1,\ Y=0), & P(X=1,\ Y=1) \end{vmatrix} = \quad \text{(adding the bottom row to the top one and the second column to the first)}$$

$$= \begin{vmatrix} 1 & P(Y=1) \\ P(X=1) & P(X=1,\ Y=1) \end{vmatrix} =$$

$$= P(X=1,\ Y=1) - P(X=1)P(Y=1).$$

$$PQD(X,Y) \Rightarrow P(X=1,\ Y=1) - P(X=1)P(Y=1) \geq 0 \Rightarrow D \geq 0 \Rightarrow TP_2(X,Y).$$

Appendices

Appendix 1.

Details of the proof of Hoeffding's lemma (Lemma 3, Part 1).

Define:

$$I(u,x) = 1 \quad \text{if} \quad u < x$$
$$= 0 \quad \text{if} \quad u \geq x.$$

Auxiliary lemma.

1. If X is a random variable then $E\, I(u,X) = P(X \geq u)$.

Pf: $E\, I(u,X) = 1 \cdot P(I(u,X) = 1) + 0 \cdot P(I(u,X) = 0)$

$= 1 \cdot P(X > u) + 0 = P(X > u)$.

Auxiliary lemma.

2. Let X_1, X_2, Y_1, Y_2 be random variables defined on the same probability space (Ω, F, P). Then,

$$(X_1 - X_2)(Y_1 - Y_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (I(u,X_1) - I(u,X_2))(I(v,Y_1) - I(v,Y_2))dudv.$$

Proof:

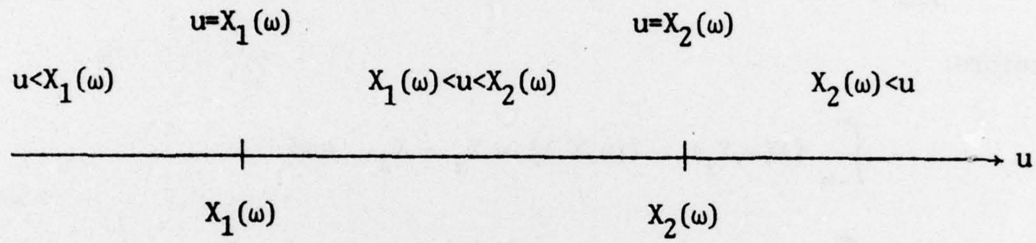
Both sides of the equality are random variables, and the double integral is interpreted as a two dimensional Lebesgue integral. The integrability of the integrand on the right hand double integral is justified by Fubini's theorem. This theorem assures that if

$\int_{-\infty}^{\infty} I(u,X_1) - I(u,X_2)du$ and $\int_{-\infty}^{\infty} (I(v,Y_1) - I(v,Y_2))dv$ exist so does

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [I(u,X_1) - I(v,X_2)][I(v,Y_1) - I(v,Y_2)]dudv.$$

Now $\int_{-\infty}^{\infty} \left\{I(u, X_1^{(\omega)}) - I\left(u, X_2^{(\omega)}\right)\right\}du$ is proved to be equal to $X_1(\omega) - X_2(\omega)$ for each $\omega\epsilon\Omega$.

Consider the case $X_1(\omega) < X_2(\omega)$.



In this case:

$$\begin{array}{lll}
I\left(u, X_1(\omega)\right) - I\left(u, X_2(\omega)\right) = 1 - 1 = 0 & \text{if} & u < X_1(\omega) \\
I\left(u, X_1(\omega)\right) - I\left(u, X_2(\omega)\right) = 1 - 1 = 0 & \text{if} & u = X_1(\omega) \\
I\left(u, X_1(\omega)\right) - I\left(u, X_2(\omega)\right) = 0 - 1 = -1 & \text{if} & X_1(\omega) < u < X_2(\omega) \\
I\left(u, X_1(\omega)\right) - I\left(u, X_2(\omega)\right) = 0 - 1 = -1 & \text{if} & u = X_2(\omega) \\
I\left(u, X_1(\omega)\right) - I\left(u, X_2(\omega)\right) = 0 - 0 = 0 & \text{if} & X_2(\omega) < u.
\end{array}$$

We thus have only to consider the integral

$\int_{X_2^+(\omega)}^{X_2^-(\omega)} \left\{I\left(u, X_1(\omega)\right) - I\left(u, X_2(\omega)\right)\right\}du$. This integral in interpreted as:

$\lim\limits_{\substack{\epsilon_1\to 0\\ \epsilon_2\to 0}} \int_{X_1(\omega)+\epsilon_1}^{X_2(\omega)-\epsilon_2} [I\left(u, X_1(\omega)\right) - I\left(u, X_2(\omega)\right)]du$ with $\epsilon_1,\epsilon_2 > 0$. (We may handle this integral as a Riemann integral, since Riemann sense and Lebesgue sense coincide in this case for the function $I\left(u, X_1(\omega)\right) - I\left(u, X_2(\omega)\right)$ on

the set $\big(X_1(\omega),\ X_2(\omega)\big)$.) The last integral equals

$$\lim_{\substack{\varepsilon_1 \downarrow 0 \\ \varepsilon_2 \downarrow 0}} \ (-1)\ X_2(\omega) - \varepsilon_2 - \big(X_1(\omega) + \varepsilon_1\big) \ = (-1)\big(X_2(\omega) - X_1(\omega)\big) = X_1(\omega) - X_2(\omega).$$

In a similar manner

$$\int_{-\infty}^{\infty} \Big[I\big(v,\ Y_1(\omega)\big) - I\big(v,\ Y_2(\omega)\big)\Big] dv = Y_1(\omega) - Y_2(\omega).$$

Therefore:

$$\int_{-\infty}^{\infty} \big(I(u,X_1) - I(u,X_2)\big) = X_1 - X_2 \quad \text{and}$$

$$\int_{-\infty}^{\infty} \big(I(v,Y_1) - I(v,Y_2)\big) = Y_1 - Y_2 \ .$$

So,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(I(u,X_1) - I(u,X_2)\big) \cdot \big(I(v,Y_1) - I(v,Y_2)\big) du dv$$

(by Fubini's theorem) as

$$= \int_{-\infty}^{\infty} I(u,X_1) - I(u,X_2) du \int_{-\infty}^{\infty} \big(I(v,Y_1) - I(v,Y_2)\big) dv$$

$$= (X_1 - X_2)(Y_1 - Y_2).$$

Appendix 2.

Theorem (Jensen (1971))

Let $\mu_2(x,y)$ be a two-dimensional distribution function which admits the series expansion

$$d\mu_2(x,y) = [1 + \sum_{r=1}^{\infty} a_r \phi_x(x)\phi_r(y)]d\mu_1(x)\mu_1(y)$$

in the functions $\{\phi_r(\cdot)\}$ and $\mu_1(\cdot)$. Then a sufficient condition that the equality

$$\int_{A\times A} d\mu_2(x,y) \geq \int_A d\mu_1(x) \int_A d\mu_1(y)$$

holds for every measurable set A is that the sequence $\{a_r\}$ be non-negative.

Remark: Note that both marginals of $\mu_2(\cdot\cdot)$ are $\mu_1(y)$. If all $a_r = 0$, $r = 1,\ldots,$ we obtain the independent case.

Proof:
$$\begin{aligned}\int_{A\times A} d\mu_2(x,y) &= \int_A d\mu_1(x) \int_A d\mu_1(y) + \\ &+ \sum_{r=1}^{\infty} a_r \int_A \phi_r(x)d\mu_1(x) \int_A \phi_r(y)d\mu_1(y) \\ &= \int_A d\mu_1(x) \int_A d\mu_1(y) + \sum_{r=1}^{\infty} a_r \left(\int_A \phi_r(x)d\mu_1(x)\right)^2 \\ &\geq \int_A d\mu_1(x) \int_A d\mu_1(y).\end{aligned}$$ □

Example: In the bivariate normal the well known expansion is:

$$d\mu_2(x,y) = [1 + \sum_{r=1}^{\infty} \rho^r \phi_r(x)\phi_r(y)]d\mu_1(x)d\mu_1(y)$$

where $\phi_r(x)$ are the Hermite polynomials and ρ is the correlation coefficient.

<u>Appendix 3</u>.

<u>Theorem 1</u> (Shannon (1948))

The function $h(\underset{\sim}{p}) = -\sum_{i=1}^{N} p_i \log p_i \le \log N$ with the equality iff $p_i = \frac{1}{N}$ $i = 1,2,\ldots N$.

<u>Lemma 1</u>. Let $p_1, p_2, \ldots, p_N$, $q_1, q_2, \ldots, q_N$ be arbitrary positive numbers with $\sum_{i=1}^{N} p_i = 1$ and $\sum_{i=1}^{N} q_j = 1$. Then,

$$-\sum_{i=1}^{N} p_i \log p_i \le -\sum_{i=1}^{N} p_i \log q_i$$

with the equality iff $p_i = 1_i$, $i = 1,\ldots,N$.

<u>Proof</u>: Consider the function $y = \log x$.

Elementary observations show that

$$\log x \le x - 1 \quad \text{for all} \quad x > 0.$$

Hence $\qquad \log q_i/p_i \le q_i/p_i - 1, \quad i = 1,2,\ldots N.$

Therefore, $\qquad p_i \log q_i - p_i \log p_i \le q_i - p_i$.

Summing up, $\qquad \sum_{i=1}^{N} p_i \log q_i - \sum_{i=1}^{N} p_i \log p_i \le 0$

or $\qquad -\sum_{i=1}^{N} p_i \log p_i \le -\sum_{i=1}^{N} p_i \log q_i$.

The equality will hold iff $\log q_i/p_i = \frac{q_i}{p_i} - 1$ for all i, i.e. iff $q_i = p_i \ \forall\, i = 1,2,\ldots N$.

Applying this lemma with $q_i = \frac{1}{N}$, we obtain

$$-\sum_{i=1}^{N} p_i \log p_i \le -\sum_{i=1}^{N} p_i \log \frac{1}{N} = N$$

with equality iff $p_i = q_i = \frac{1}{N}$, $i = 1,2,\ldots,N$. □

<u>Theorem 2</u> Let X and Y be discrete r.v.'s with joint entropy $-\sum_{i=1}^{n,m} p_{ij} \log p_{ij}$ and marginal entropies $-\sum_{i=1}^{n} p_i \log p_i$, $-\sum_{i=1}^{m} q_i \log q_i$, where $P(X=x_i, Y=y_j) = p_{ij}$; $P(X=x_i) = p_i$ and $P(Y=y_j) = q_j$, $\begin{matrix} i = 1,2,\ldots,n \\ j = 1,2,\ldots,m. \end{matrix}$

Then:

$$-\sum_{i,j}^{n,m} p_{ij} \log p_{ij} \le -\sum_{i,j}^{n,m} p_i q_j \log p_i q_j \,. \qquad \bullet$$

<u>Proof</u>:

Let $$h(\underset{\sim}{p}) = -\sum_{i=1}^{n} p_i \log p_i = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log p_i$$

$$h(\underset{\sim}{q}) = -\sum_{j=1}^{m} q_j \log q_j = -\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij} \log q_j \,.$$

$$h(\underset{\sim}{p}) + h(\underset{\sim}{q}) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log p_i q_j \ge -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log p_{ij} \,.$$

However, $$-\sum_{i,j}^{n,m} p_i q_j \log p_i q_j = -\sum_{i,j}^{n,m} p_i q_j (\log p_i + \log q_j) = h(\underset{\sim}{p}) + h(\underset{\sim}{q}).$$

(Equality is attained in $\bullet$ iff $p_{ij} = p_i q_j$, $\begin{array}{l} i = 1,\ldots n \\ j = 1,\ldots m \end{array}$.)

Appendix 4.

Continuous and multivariate extensions of the information-theoretic measure of dependence.

Let (X,Y) be a random vector with joint density $f(x,y)$, marginals $f(x) = \int f(x,y)dy$ and $g(y) = \int f(x,y)dx$. Assume that $f(x,y) > 0$ a.e. with respect to two-dimensional Lebesgue measure, $f(x) > 0$, $g(y) > 0$ a.e. with respect to Lebesgue measure and

$$E \log f(X,Y), \quad E \log f(X), \quad E \log g(Y)$$

exist.

Lemma 1. $-\int g(y)\log g(y)dy \le -\int g(y)\log f(y)dy$

or $$- E \log g(Y) \le - E \log f(Y)$$

with equality iff $g(x) = f(x)$ a.e. with respect to Lebesgue measure.

Proof: From the basic inequality

$$\log x \le x - 1 \quad \text{for all} \quad x > 0$$

we have

$$\log \frac{f(y)}{g(y)} \le \frac{f(y)}{g(y)} - 1$$

or $$\log f(y) - \log g(y) \le \frac{f(y)}{g(y)} - 1$$

or $$g(y)\log f(y) - g(y)\log g(y) \le f(y) - g(y),$$

hence: $$-\int g(y)\log g(y)dy \le -\int g(y)\log f(y)dy$$

(since $\int f(y)dy = \int g(y)dy = 1$). The equality part follows from the fact that $\log x = x - 1$ only iff $x = 1$.

Lemma 2. (Bivariate version of Lemma 1):

$$-\int\int f(x,y)\log f(x,y)dxdy \leq -\int\int f(x,y)\log f(x)g(y)dxdy$$

with equality iff $f(x,y) = f(x)g(y)$ a.e. w.r. to Lebesgue measure on R^2.

Proof: Using Lemma 1:

$$\log \frac{f(x)g(y)}{f(x,y)} \leq \frac{f(x)g(y)}{f(x,y)} - 1$$

or $$f(x,y)\log f(x)g(y) - f(x,y)\log f(x,y) \leq f(x)g(y) - f(x,y)$$

or $$\iint f(x,y)\log f(x)g(y)dxdy - \iint f(x,y)\log f(x,y)dxdy \leq 0$$

or $$-\iint f(x,y)\log f(x,y)dxdy \leq -\iint f(x,y)\log f(x)g(y)dxdy. \qquad \square$$

(The equality part follows from the basic inequality in Lemma 1.)

Remark. Lemma 2 is a particular case of a more general result.

In fact we can prove that:

$$-\iint f(x,y)\log f(x,y)dxdy \leq -\iint f(x,y)\log h(x,y)dxdy$$

where $h(x,y)$ is another bivariate density with the same support as $f(x,y)$.

This can be derived by considering the inequality

$$\log \frac{h(x,y)}{f(x,y)} \leq \frac{h(x,y)}{f(x,y)} - 1.$$

We are now ready to prove.

Theorem 1. Under the assumptions above,

$$-\iint f(x,y)\log f(x,y)dxdy \leq -\iint f(x)g(y)\log f(x)g(y)dxdy,$$

with equality iff $f(x,y) = f(x)g(y)$ a.e. w.r. to Lebesgue measure on R^2, or X, Y independent.

$$\text{RHS} = -\iint f(x)g(y)[\log f(x) + \log g(y)]dxdy$$

$$= -\int f(x)\log f(x)dx - \int g(y)\log g(y)dy$$

$$= -\iint f(x,y)\log f(x)dxdy - \iint f(x,y)\log g(y)dydx$$

$$= -\iint f(x,y)\log f(x)g(y)dxdy.$$

The interchange of integral signs is justified by Fubini's theorem (in the second double integral) since

$$\iint |f(x,y)\log g(y)|dxdy = \iint |f(x,y)|dx|\log g(y)|dy$$

$$= \int f(y)|\log g(y)|dy = E|\log g(X)| < \infty$$

by assumption.

Now applying Lemma 2:

$$-\iint f(x,y)\log f(x,y)dxdy \le -\iint f(x,y)\log f(x)g(y)dxdy. \qquad \square$$

Corollary 1. In a denumerable discrete case, we have

$$-\sum_{i,j} p_{ij} \log p_{ij} \le -\sum_{i,j} p_i q_j \log p_i q_j$$

where $p_i = \sum_{j=1}^{\infty} p_{ij}$, $q_j = \sum_{i=1}^{\infty} p_{ij}$.

Remarks: A denumerable discrete version of Lemma 2 is:

$$-\sum_{i,j} p_{ij} \log p_{ij} \le -\sum_{i,j} p_{ij} \log p_i q_j ,$$

where $p_i = \sum_{j=1}^{\infty} p_{ij}$ and $q_j = \sum_{i=1}^{\infty} p_{ij}$.

A special case of the Remark to Lemma 2 is:

$$-\sum_{i,j} p_{ij} \log p_{ij} \le -\sum_{i,j} p_{ij} \log h_{ij} ,$$

where h_{ij} is another bivariate discrete distribution with some probability support as p_{ij} .

Multivariate extensions of the above cases can be summarized by the following two theorems. The first is an extension of Lemma 2, while the second is that of Theorem 1.

Theorem 2.

Let $(\underset{\sim}{X},\underset{\sim}{Y})$ be on m+n dimensional vector.

Then $\int_{m+n}\!\!\cdots\!\int f(\underset{\sim}{x},\underset{\sim}{y}) \log f(\underset{\sim}{x},\underset{\sim}{y}) d\underset{\sim}{x} d\underset{\sim}{y} \le -\int_{m+n}\!\!\cdots\!\int f(\underset{\sim}{x},\underset{\sim}{y}) \log f(\underset{\sim}{x}) g(\underset{\sim}{y}) d\underset{\sim}{x} d\underset{\sim}{y}$

or $$-E \log f(\underset{\sim}{X},\underset{\sim}{Y}) \le -E \log f(\underset{\sim}{X}) g(\underset{\sim}{Y}).$$

(With equality iff X, Y are independent.)

Theorem 3.

$$-\int_{m+n}\!\!\cdots\!\int f(\underset{\sim}{x},\underset{\sim}{y}) \log f(\underset{\sim}{x},\underset{\sim}{y}) d\underset{\sim}{x} d\underset{\sim}{y} \le -\int_{m+n}\!\!\cdots\!\int f(\underset{\sim}{x}) g(\underset{\sim}{y}) \log f(\underset{\sim}{x}) g(\underset{\sim}{y}) d\underset{\sim}{x} d\underset{\sim}{y}.$$

(With equality iff X, Y are independent.)

Appendix 5

Theorem. For the bivariate FGM (Farlie-Gumbel-Morgenstern) family of bivariate distributions given by

$$G_{\chi_1,\chi_2}(x_1,x_2) = G_{\chi_1}(x_1) G_{\chi_2}(x_2)[1 + \alpha(1 - G_{\chi_1}(x_1))(1 - G_{\chi_2}(x_2))] \quad ; \quad |\alpha| < 1$$

or equivalently

$$F_{X_1,X_2}(x_1,x_2) = F_{X_1}(x_1)F_{X_2}(x_2)[1 + \alpha(1 - F_{X_1}(x_1))(1 - F_{X_2}(x_2))],$$

the G-orthant positive dependence is equivalent to the property

$$h_{X_j}(x_j) > h_{\underset{\sim}{X}}(\underset{\sim}{x})_j \; .$$

<u>Proof</u>: For this family $R_G(x_1,x_2) \geq 1$ iff $\alpha \geq 0$.

Moreover for this family:

$$\begin{aligned} h_{X_1,X_2}(\underset{\sim}{x})_j &= h_{X_j}(x_j) - \alpha\{1 - G_{X_{3-j}}(x_{3-j})\}f_{X_j}(x_j)/[1 + \alpha F_{X_1}(x_1)F_{X_2}(x_2)] \\ &= [1 - \{\beta(G_{X_j}(x_j))^{-1} - 1\}^{-1}]h_{X_j}(x_j), \end{aligned}$$

where $\beta = 1 + [\alpha\, F_{X_{3-j}}(x_{3-j})]^{-1}$, $j = 1,2$.

Since β has the same sign as α, $h_{X_j}(x_j) - h_{X_1,X_2}(\underset{\sim}{x})_j$ has the same sign α; thus $h_{X_j}(x_j) > h_{\underset{\sim}{X}}(x)_j \iff \alpha > 0 \iff R_G(x_1,x_2) \geq 1$. □

Selected References

Barlow, R.E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Kinston, N.Y.

Esary, J.D. and Proschan, F. (1972). Relationship among some concepts of bivariate dependence, *Ann. Math. Statist.*, 43, 541-

Esary, K.D., Proschan, F. and Walkup, D.W. (1967). Association of random variables with applications, *Ann. Math. Statist.*, 38, 1466-74.

Hoeffding, W. (1940). Masstabinvariante korrelations-theory, *Schriften Math. Insti. Univ. Berlin*, 5, 181-233.

Jensen, D.R. (1971). A note on positive dependence and the structure of bivariate distributions, *SIAM Journ. Appl. Math.*, 20, No. 4, 749-753.

Johnson, N.L. and Kotz, S. (1975). On some generalized FGM distributions. *Commun. in Statist.*, 4, 415-427.

Johnson, N.L. and Kotz, S. (1975). A vector-valued multivariate hazard rate, *Journ. Multiv. Analysis.*, 5, 53-66.

Karlin, S. (1968). *Total Positivity I*. Academic Press.

Lehmann, E.L. (1966). Some concepts of dependence, *Ann. Math. Statist.*, 37, 1137-1153.

Linfoot, E.H. (1957). An informational measure of correlation, *Information and Control*, 1, 85-89.

Renyi, A. (1959). New version of the probabilistic generalization of the large sieve, *Acta Math. Acad. Sci. Hungar.*, 10, 217-226.

Schweizer, B. and Wolff, E.F. (1976). Sur une measure de dependance pour les variables aleatoires, *C. R. Acad. Sci., Paris*, Ser. A, 283, 609-611.

Shannon, C.E. (1948). A mathematical theory of communication, *Bell System Tech. J.*, 24, 379, 623.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE Rev

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (19) 18 AFOSR-TR-77-0745-[redacted]

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER (9) Revised Edition

4. TITLE (and Subtitle) (6) ON MEASURES OF DEPENDENCE. A SURVEY OF RECENT DEVELOPMENTS. Revision,

5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s) (10) S. Kotz [redacted] C. Soong

8. CONTRACT OR GRANT NUMBER(s) (15) AFOSR-75-2837

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Temple University
Department of Mathematics
Philadelphia, PA 19122

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) (17) 61102F 2304 A5

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Office of Scientific Research/NM
Bolling AFB, Washington, DC 20332

12. REPORT DATE (11) July 1977

13. NUMBER OF PAGES ~~50~~ (12) 52p

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Supersedes AD-A041814

mc

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

bivariate distributions; FGM distributions; multivariate dependence; Hoeffding's lemma; multivariate hazard rates; correlation ratio; cou;as; association; information theory; reliability

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The purpose of these notes is to survey the recent literature on dependence between two random variables and to stimulate research which will extend (some of) these concepts and relations to the case of 3 or more variables.

404207

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)